

WHAT IS CLAIMED IS:

1. A system for collecting identifiers for updating a filtering system which controls access to Internet websites/pages between a local area network and an Internet, comprising:

a workstation configured for a user to send an identifier to request an Internet website/page;

an Internet gateway system coupled to the workstation and configured to receive the identifier and to allow or deny access to the Internet website/page associated with the identifier;

a master database of identifiers along with one or more categories associated with each identifier;

a filter system coupled to the Internet gateway system and configured to receive the identifier from the Internet gateway system, determine whether the identifier is in the master database, send the identifier to a database factory if the identifier is not in the master database, and apply one or more rules to one or more categories that are associated with the identifier, wherein the one or more categories are received from the database factory; and

a database factory configured to receive the identifier from the filter system if the identifier was not in the master database, determine whether the identifier was previously categorized by the database factory, if the identifier was not previously categorized, determine the one or more categories to associate with the identifier and provide the one or more categories to the filter system, else provide the one or more categories that were previously associated with the identifier.

2. The system of Claim 1, wherein the identifier is in the master database and is associated with the one or more categories.

3. The system of Claim 1, wherein the filter system is further configured to pre-filter the identifier and/or the Internet website/page associated with the identifier for a data characteristic that is indicative of the one or more categories, and associating one or more indicators with the identifier.

4. The system of Claim 3, wherein pre-filtering can be performed on text strings, graphics, and audio that are associated with the identifier and/or the Internet website/page.

5. The system of Claim 3, wherein the one or more indicators can include a category flag.

6. The system of Claim 5, wherein the filter system uses the one or more indicators to screen the identifier prior to sending the identifier to the database factory.

7. The system of Claim 1, wherein the Internet gateway system comprises:  
a firewall module configured to provide an electronic boundary between the workstation and the Internet; and  
a router module configured to find a best path from the firewall module to the Internet website/page associated with the identifier.

8. The system of Claim 7, wherein the filter system comprises:  
a management module configured to provide an interface for a system administrator to select the one or more rules that are applied to the one or more categories associated with the identifier;  
an uncategorized database configured to store the identifier if the identifier is not in the master database; and  
an upload/download manager module configured to send the stored identifier to the database factory and to receive the one or more categories from the database factory.

9. The system of Claim 8, wherein the uncategorized database includes a request frequency that is associated with the identifier and indicates the frequency of requests for the identifier in the uncategorized database.

10. The system of Claim 9, wherein the upload/download manager module is configured to send the request frequency from the uncategorized database to the database factory.

11. The system of Claim 10, wherein the master database includes a request frequency that is associated with the identifier and indicates the frequency of requests for the identifier in the master database.

12. The system of Claim 11, wherein the upload/download manager module is configured to send the request frequency from the master database to the database factory.

13. The system of Claim 8, wherein the one or more rules include blocking access to the Internet website/page based on the one or categories associated with the identifier and the user.

14. The system of Claim 8, wherein the one or more rules include allowing access to the Internet website/page based on the one or categories associated with the identifier and the user.

15. The system of Claim 1, wherein the database factory comprises:

an upload/download module configured to receive the identifier from the filter system and provide the one or more categories to the filter system;

an identifier processing module configured to receive the identifier from the upload/download module and determine whether the identifier has been previously categorized by the database factory;

a categorization system module configured to categorize the identifier if not previously categorized by the database factory; and

a database of categorized identifiers configured to store the identifier and the one or more categories.

16. The system of Claim 15, wherein the upload/download module is configured to receive a request frequency from the filter system to prioritize the identifier.

17. The system of Claim 16, wherein the request frequency is associated with the identifier and indicates the frequency of requests for the identifier in the uncategorized database.

18. The system of Claim 16, wherein the request frequency is associated with the identifier and indicates the frequency of requests for the identifier in the master database.

19. The system of Claim 15, wherein the identifier is a uniform resource locator (URL).

20. The system of Claim 15, wherein the identifier is an Internet Protocol (I.P.) address.

21. The system of Claim 15, wherein the identifier is a domain name.

22. The system of Claim 15, further comprising:

a second filter system; and

a second Internet gateway system coupled to the second filter system and the database factory.

23. The system of Claim 22, wherein the identifier processing module is further configured to merge and sort the identifier and a second identifier received from the second Internet gateway system.

24. A method for adapting a filter system which controls access to Internet sites, the method comprising:

receiving a request from a user in the form of an identifier to access a website/page;

determining whether the identifier is in a master database of categorized identifiers and one or more categories associated with the identifier;

if the identifier is not in the master database, determining whether the identifier is in an uncategorized database, else applying one or more rules to the one or more categories associated with the identifier;

if the identifier is not in the uncategorized database, posting the identifier to the uncategorized database, else updating an uncategorized database request frequency in the uncategorized database that is associated with the identifier;

uploading the uncategorized database to a database factory;

determining whether each identifier has been previously categorized by the database factory;

for each identifier that was not previously categorized, categorizing each identifier and/or a website/page associated with the identifier to select one or more categories to associated with each identifier; and

posting each identifier along with its selected one or more categories into a database of categorized sites; and

downloading the database of categorized sites to the filter system for incorporation into the master database.

25. The method of Claim 24, further comprising:

updating a master database request frequency in the master database if the identifier is in the master database; and

uploading the master database request frequency and the associated identifier to the database factory.

26. The method of Claim 24, wherein the one or more rules include blocking access to the website/page based on the one or more categories associated with the identifier and the user.

27. The method of Claim 24, wherein the one or more rules include allowing access to the website/page based on the one or more categories associated with the identifier and the user.

28. The method of Claim 24, wherein the uncategorized database further includes additional data associated with the identifier.

29. The method of Claim 28, wherein the additional data includes a request frequency.

30. The method of Claim 28, wherein the additional data includes an indicator.

31. The method of Claim 28, wherein the additional data includes a trace ID.

32. The method of Claim 28, wherein the additional data includes a primary language used by the filter system.

33. The method of Claim 28, further comprising merging and sorting the uncategorized database based on each identifier and the additional data associated with each identifier.

34. The method of Claim 28, further comprising:

pre-filtering the identifier and/or website/page associated with the identifier for data characteristics that are indicative of the one or more categories; and associating one or more indicators with the identifier.

35. The method of Claim 34, wherein the pre-filtering is performed on text strings, graphics, and audio that are associated with the identifier and or website/page associated with the identifier.

36. The method of Claim 34, wherein the one or more indicators can include a category flag.

37. The method of Claim 36, further comprising screening the identifier using the one or more indicators prior to sending the identifier to the database factory.

38. The method of Claim 37, wherein the identifier is a uniform resource locator (URL).

39. The method of Claim 37, wherein the identifier is an Internet Protocol (I.P.) address.

40. The method of Claim 37, wherein the identifier is a domain name.

41. A method for collecting collection data for updating a filtering system which controls access to Internet sites, the method comprising:

receiving a request from a user in the form of an identifier to access a website/page;

determining whether the identifier is stored in a master database of categorized identifiers;

if the identifier is stored, recalling a category that is associated with the identifier and applying a rule to the identifier that is associated with the category;

if the identifier is not stored, determining whether the identifier is stored in an uncategorized database; and

if the identifier is not stored in the uncategorized database, posting the identifier to the uncategorized database, else updating an uncategorized database request frequency associated with the identifier.

42. The method of Claim 41, further comprising:

updating a master database request frequency in the master database if the identifier is stored in the master database.

43. The method of Claim 41, further comprising:

pre-filtering the identifier and/or website/page associated with the identifier for data characteristics that are indicative of one or more categories;

determining whether data characteristics were found during the pre-filtering; and

if data characteristics were found, associating one or more indicators with the identifier in the uncategorized database.

44. The method of Claim 43, wherein the rule blocks access to the website/page based on the category associated with the identifier and the user.

45. The method of Claim 43, wherein the rule allows access to the website/page based on the category associated with the identifier and the user.

46. The method of Claim 43, wherein the identifier is a uniform resource locator (URL).

47. The method of Claim 43, wherein the identifier is an Internet Protocol (I.P.) address.

48. The method of Claim 43, wherein the identifier is a domain name.

49. The method of Claim 43, wherein the uncategorized database further includes additional data associated with the identifier.

50. The method of Claim 49, wherein the additional data includes a trace ID.

51. The method of Claim 49, wherein the additional data includes a primary language used by the filter system.

52. The method of Claim 49, wherein the additional data includes the uncategorized database request frequency.

53. The method of Claim 43, wherein the master database further includes additional data associated with the identifier.

54. The method of Claim 53, wherein the master database further includes the master database request frequency.

55. A method for processing and uploading identifiers for updating a filtering system which controls access to Internet sites, the method comprising:

requesting a download of identifiers and their associated categories from a database factory;

determining whether a database of identifiers are to be uploaded to the database factory;

if the database of identifiers are to be uploaded to the database factory, retrieving identifiers from the database of identifiers; and

uploading the database of identifiers to the database factory.

56. The method of Claim 55, wherein the database of identifiers includes an uncategorized database of identifiers.

57. The method of Claim 55, wherein the database of identifiers includes a master database of identifiers.

58. The method of Claim 57, wherein the master database of identifiers includes a master database request frequency which indicates a frequency of requests for identifiers in the master database of identifiers.

59. The method of Claim 56, further comprising:

determining whether one or more indicators, which relate to one or more categories, are associated with the uncategorized identifiers; and

referring to the one or more indicators to prioritize the uncategorized identifiers prior to uploading the uncategorized database of identifiers to the database factory.

60. The method of Claim 59, further comprising uploading additional data associated with the uncategorized identifiers to the database factory.

61. The method of Claim 60, wherein the additional data includes a request frequency.

62. The method of Claim 60, wherein the additional data includes an indicator.

63. The method of Claim 60, wherein the additional data includes a trace ID.

64. The method of Claim 60, wherein the additional data includes a primary language used by the filter system.

65. The method of Claim 60, further comprising processing the uncategorized identifier and the additional data prior to uploading to the database factory.

66. The method of Claim 65, wherein the processing comprises:

formatting the uncategorized identifiers and the additional data using a markup language; and

limiting the size of an upload file which includes the uncategorized identifiers and the additional data.

67. The method of Claim 66, further comprising:

encrypting the uncategorized identifiers and the additional data; and



compressing the uncategorized identifiers and the additional data.

68. The method of Claim 67, wherein encrypting is performed using a data encryption standard (DES).

69. The method of Claim 67, wherein the uncategorized identifiers are uniform resource locators (URLs).

70. The method of Claim 67, wherein the uncategorized identifiers are Internet Protocol (I.P.) addresses.

71. The method of Claim 67, wherein the uncategorized identifiers are domain names.

72. The method of Claim 55, wherein uploading the database of identifiers is periodic.

73. The method of Claim 55, wherein uploading the database of identifiers is random.

74. The method of Claim 55, wherein uploading the database of identifiers is at a set time.

75. The method of Claim 55, wherein uploading the database of identifiers is in response to polling by the database factory.

76. A method for processing identifiers for updating a filtering system which controls access to Internet sites, the method comprising:

receiving identifiers at a database factory from a filter system;

determining whether each of the identifiers has been previously categorized;

for each identifier that was not previously categorized by the database factory, categorizing each of the identifiers, a website/page associated with the identifier, and/or additional data associated with the identifier, and associating output with each identifier; and

posting the identifier along with the associated output into a database of categorized sites.

77. The method of Claim 76, wherein the output includes one or more categories.

78. The method of Claim 76, wherein the output includes a statistical probability.

79. The method of Claim 76, wherein the output includes a multidimensional vector.

80. The method of Claim 76, wherein categorizing each of the identifiers is performed by an automated classification engine.

81. The method of Claim 76, wherein the identifiers include an uncategorized identifier.

82. The method of Claim 76, wherein the identifiers include a categorized identifier.

83. The method of Claim 82, further comprising receiving at the database factory a master database request frequency which indicates the frequency that the categorized identifier was requested at the filter system.

84. The method of Claim 76, further comprising merging and sorting the identifier and a second identifier received from a second filter system.

85. The method of Claim 80, further comprising determining whether processing of the output associated with the uncategorized identifier is performed.

86. The method of Claim 85, wherein the processing of the output includes automatically posts the uncategorized identifier to the master database.

87. The method of Claim 85, wherein the processing of the output includes verifying that the output associated with the uncategorized identifier is correct.

88. The method of Claim 87, wherein verifying is performed by human review.

89. The method of Claim 88, further comprising providing feedback from the human review to the automated classification engine.

90. The method of Claim 76, further comprising processing the identifiers and the additional data.

91. The method of Claim 90, wherein processing includes reassembling the identifiers and the additional data using a markup language.

92. The method of Claim 91, further comprising:  
decrypting the identifiers and the additional data; and  
decompressing the identifiers and the additional data.

93. The method of Claim 92, wherein decrypting is performed using a data encryption standard (DES).

94. The method of Claim 76, wherein the identifier is a uniform resource locator (URL).

95. The method of Claim 76, wherein the identifier is an Internet Protocol (I.P.) address.

96. The method of Claim 76, wherein the identifier is a domain name.

97. A system for collecting identifiers for updating a filtering system which controls access to a wide area network (WAN) of websites/pages, comprising:

a master database including one or more identifiers, and one or more categories associated with each of the one or more identifiers;

an access system coupled to the WAN and configured to send an identifier request if the identifier request is not in the master database; and

a database factory configured to receive the identifier request, select one or more categories to associate with the identifier request if the one or more categories were not previously associated with the identifier, and provide the selected one or more categories to the master database.

98. The system of Claim 97, wherein the access system comprises an Internet gateway system coupled to a filter system.

99. The system of Claim 98, wherein the identifier is a uniform resource locator (URL).

100. The system of Claim 98, wherein the identifier is an Internet Protocol (I.P.) address.

101. The system of Claim 98, wherein the identifier is a domain name.

102. The system of Claim 98, wherein the master database includes a request frequency that is associated with the identifier and indicative of the frequency of requests for the identifier at the filter system.

103. The system of Claim 102, wherein the database factory is configured to receive the request frequency to prioritize the received identifiers.